# SYSTEMS AND METHODS FOR AIDING HUMAN TRANSLATION

## RELATED APPLICATIONS

**[0001]** This application claims priority under 35 U.S.C. § 119 based on U.S. Provisional Application Nos. 60/394,064 and 60/394,082, filed July 3, 2002, and Provisional Application No. 60/419,214, filed October 17, 2002, the disclosures of which are incorporated herein by reference.

**[0002]** This application is related to U.S. Patent Application, Serial No. _____ (Docket No. 02-4036), entitled, "Systems and Methods for Facilitating Playback of Media," filed concurrently herewith and incorporated herein by reference.

## GOVERNMENT CONTRACT

**[0003]** The U.S. Government may have a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of Contract No. 1999*S018900*000 awarded by the Federal Broadcast Information Service.

## BACKGROUND OF THE INVENTION

### Field of the Invention

**[0004]** The present invention relates generally to language translation and, more particularly, to systems and methods for aiding a human in translating audio data.

### Description of Related Art

**[0005]** There are three major tasks when performing translations of an audio signal: selection, translation, and publication. During selection, a human translator chooses a segment of audio to translate. During translation, the translator actually translates the audio segment. During publication, the translator publishes or saves the translation results.

**[0006]** Human translation is a slow and time-consuming process. As a result, the human translator typically translates only important segments of an audio signal. The translator will often work from a recorded audio signal to skim the complete audio signal, listening for segments that are suitable for translation. The translator then replays selected segments, translating the speech while transcribing them with a word processor. To accurately transcribe the audio segments, the translator will usually go through the audio segments many times, rewinding the audio repeatedly to keep the translation synchronized with the playback. Only after the translator feels that the translated audio segment is accurate and complete will the translator publish the translation results.

**[0007]** As a result, there is a need for mechanisms that facilitate and expedite the translation of an audio signal.

## SUMMARY OF THE INVENTION

**[0008]** Systems and methods consistent with the present invention address this and other needs by providing a transcription of an audio signal, along with the original audio signal, to a translator to assist the translator in translating the audio signal. The systems and methods visually synchronize the playback of the audio signal with the transcription to aid the translation process.

**[0009]** In one aspect consistent with the principles of the invention, a system aids a user in translating an audio signal that includes speech from one language to another. The system retrieves a textual representation of the audio signal and presents the textual representation to the user. The system receives selection of a segment of the textual representation for translation and obtains a portion of the audio signal corresponding to the segment. The system then provides the segment of the textual representation and the portion of the audio signal to the user to help the user translate the audio signal.

**[0010]** According to another aspect of the invention, a graphical user interface is provided. The graphical user interface includes a transcription section, a translation section, and a play button. The transcription section includes a transcription of non-text information in a first language. The translation section receives a translation of the non-text information into a second language. The play button, when selected, causes retrieval of the non-text information to be initiated, playing of the non-text information, and the playing of the non-text information to be visually synchronized with the transcription in the transcription section.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0011]** The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate the invention and, together with the description, explain the invention. In the drawings,

**[0012]** Fig. 1 is a diagram of a system in which systems and methods consistent with the present invention may be implemented;

**[0013]** Fig. 2 is an exemplary diagram of the server of Fig. 1 according to an implementation consistent with the principles of the invention;

**[0014]** Fig. 3 is an exemplary diagram of the client of Fig. 1 according to an implementation consistent with the principles of the invention;

**[0015]** Fig. 4 is a flowchart of exemplary processing for presenting information for perusal by a human translator according to an implementation consistent with the principles of the invention;

**[0016]** Fig. 5 is a diagram of an exemplary graphical user interface that may be presented to a user according to an implementation consistent with the principles of the invention;

**[0017]** Fig. 6 is a diagram of the graphical user interface of Fig. 5 that illustrates a user's request to play back an original audio signal;

**[0018]** Fig. 7 is a diagram of the graphical user interface of Fig. 5 that illustrates synchronization of a document to the playback of the original audio signal;

**[0019]** Fig. 8 is a flowchart of exemplary processing for translating an audio signal according to an implementation consistent with the principles of the invention;

**[0020]** Fig. 9 is a diagram of an exemplary graphical user interface that may be presented to a user in an implementation consistent with the principles of the invention; and

**[0021]** Fig. 10 is a diagram of the graphical user interface of Fig. 9 that illustrates a user's translation of an audio signal.

## DETAILED DESCRIPTION

**[0022]** The following detailed description of the invention refers to the accompanying drawings. The same reference numbers in different drawings may identify the same or similar elements. Also, the following detailed description does not limit the invention. Instead, the scope of the invention is defined by the appended claims and equivalents.

**[0023]** Systems and methods consistent with the present invention aid a human translator in translating an audio stream from one language to another. The systems and methods present the human translator with the audio stream, along with a transcription of the audio stream. The systems and methods visually synchronize the playing back of the audio with the words in the transcription. As will be apparent below, such systems and methods help the human translator translate the audio stream more efficiently, quickly, and accurately.

## EXEMPLARY SYSTEM

**[0024]** Fig. 1 is a diagram of an exemplary system 100 in which systems and methods consistent with the present invention may be implemented. System 100 may include server 110, metadata database 120, database of original media 130, and clients 140 interconnected via a network 150. Network 150 may include any type of network, such as a local area network (LAN), a wide area network (WAN), a public telephone network (e.g., the Public Switched Telephone Network (PSTN)), a virtual private network (VPN), or a combination of networks. Server 110, database 130, and clients 140 may connect to network 150 via wired, wireless, and/or optical connections.

**[0025]** Generally, clients 140 may interact with server 110 to obtain information from metadata database 120. The information may include a textual representation (or transcription)

of audio data. A user of one of clients 140 may peruse the information to identify segments to be translated. Client 140 may then obtain the original audio from database of original media 130 either directly or via server 110. Client 140 may present the information and original audio to the user in such a manner that facilitates the user's translation of the audio.

**[0026]** Each of the components of system 100 will now be described in more detail.

Server 110

**[0027]** Server 110 may include a computer or another device that is capable of servicing client requests for information and providing such information to a client 140. Fig. 2 is an exemplary diagram of server 110 according to an implementation consistent with the principles of the invention. Server 110 may include bus 210, processor 220, main memory 230, read only memory (ROM) 240, storage device 250, input device 260, output device 270, and communication interface 280. Bus 210 permits communication among the components of server 110.

**[0028]** Processor 220 may include any type of conventional processor or microprocessor that interprets and executes instructions. Main memory 230 may include a random access memory (RAM) or another type of dynamic storage device that stores information and instructions for execution by processor 220. ROM 240 may include a conventional ROM device or another type of static storage device that stores static information and instructions for use by processor 220. Storage device 250 may include a magnetic and/or optical recording medium and its corresponding drive.

**[0029]** Input device 260 may include one or more conventional mechanisms that permit an operator to input information to server 110, such as a keyboard, a mouse, a pen, voice recognition

and/or biometric mechanisms, etc. Output device 270 may include one or more conventional mechanisms that output information to the operator, including a display, a printer, a pair of speakers, etc. Communication interface 280 may include any transceiver-like mechanism that enables server 110 to communicate with other devices and/or systems. For example, communication interface 280 may include mechanisms for communicating with another device or system via a network, such as network 150.

**[0030]** As will be described in detail below, server 110, consistent with the present invention, services requests for information and manages access to metadata database 120. Server 110 may perform these tasks in response to processor 220 executing sequences of instructions contained in, for example, memory 230. These instructions may be read into memory 230 from another computer-readable medium, such as storage device 250, or from another device via communication interface 280.

**[0031]** Execution of the sequences of instructions contained in memory 230 causes processor 220 to perform processes that will be described later. Alternatively, hardwired circuitry may be used in place of or in combination with software instructions to implement processes consistent with the present invention. Thus, processes performed by server 110 are not limited to any specific combination of hardware circuitry and software.

Metadata Database 120

**[0032]** Metadata database 120 may include a relational database, or another type of database, that stores metadata and other information relating to audio data in any language. An audio processing system (not shown), such as the one described in John Makhoul et al., "Speech and Language Technologies for Audio Indexing and Retrieval," Proceedings of the IEEE, Vol. 88,

No. 8, August 2000, pp. 1338-1353, may capture audio data from various sources, process the audio data, and create an automated transcription and metadata relating to the audio data.

**[0033]** For example, the media processing system may segment an input audio stream by speaker, cluster audio segments from the same speaker, identify speakers known to the system, and transcribe the spoken words. The media processing system may also segment the input stream into stories, based on their topic content, and locate the names of people, places, and organizations. The media processing system may further analyze the input stream to identify when each word is spoken. The media processing system may include any or all of this information in the transcription and metadata relating to the input stream.

Database of Original Media 130

**[0034]** Database of original media 130 may include a conventional database that stores audio (or other types of media) in any language. The audio may be processed by a known audio compression technique, such as MP3 compression, and stored in database 130. The audio stored in database 130 may correspond to the information in metadata database 120. In other words, the original audio may include the data from which the transcription and metadata was created. In other implementations, database 130 may contain additional audio, or another type of media, for which there is no corresponding information in metadata database 120.

**[0035]** The original audio may be stored in such a way that it is easily retrievable as a whole and in portions. For example, a portion of an audio signal may be retrieved by specifying that the portion of the signal that occurred between 8:05 a.m. and 8:08 a.m. is desired. The database 130 may then provide the desired audio as streaming audio to client 140, for example.

Client 140

**[0036]** Client 140 may include a personal computer, a laptop, a personal digital assistant, or another type of device that is capable of interacting with server 110 and database of original media 130 to obtain information for translation or perusal. Client 140 may present the information to a user via a graphical user interface (GUI), possibly within a web browser window or a word processing window.

**[0037]** Fig. 3 is an exemplary diagram of client 140 according to an implementation consistent with the principles of the invention. Client 140 may include a bus 310, a processor 320, a memory 330, one or more input devices 340, one or more output devices 350, and a communication interface 360. Bus 310 may permit communication among the components of client 140.

**[0038]** Processor 320 may include any type of conventional processor or microprocessor that interprets and executes instructions. Memory 330 may include a RAM or another type of dynamic storage device that stores information and instructions for execution by processor 320; a ROM or another type of static storage device that stores static information and instructions for use by processor 320; and/or some other type of magnetic or optical recording medium and its corresponding drive. For example, memory 330 may include both volatile and non-volatile memory devices.

**[0039]** Input devices 340 may include one or more conventional mechanisms that permit a user to input information into client 140 or control operation of client 140, such as a keyboard, mouse, pen, etc. In one implementation, input devices 340 may include a foot pedal that permits a user to control the playback of an audio signal. Output devices 350 may include one or more

conventional mechanisms that output information to the user, including a display, a printer, a pair of speakers, etc. Communication interface 360 may include any transceiver-like mechanism that enables client 140 to communicate with other devices and systems via a network, such as network 150.

**[0040]** As will be described in detail below, client 140, consistent with the present invention, aids a user in translating an audio signal by, for example, presenting a textual representation of the audio signal in a same window as will be used for the transcription and visually synchronizing the playing back of the audio signal with the textual representation of the audio signal. Client 140 may perform these operations in response to processor 320 executing software instructions contained in a computer-readable medium, such as memory 330.

**[0041]** The software instructions may be read into memory 330 from another computer-readable medium or from another device via communication interface 360. The software instructions contained in memory 330 causes processor 320 to perform processes that will be described later. Alternatively, hardwired circuitry may be used in place of or in combination with software instructions to implement processes consistent with the present invention. Thus, processes performed by client 140 are not limited to any specific combination of hardware circuitry and software.

## EXEMPLARY PROCESSING

**[0042]** Fig. 4 is a flowchart of exemplary processing for presenting information for perusal by a human translator according to an implementation consistent with the principles of the invention. Processing may begin with the user (i.e., human translator) inputting, into client 140, a request for information. For example, a typical request might be as specific as "give me audio

from Al Jazeera for January 3, 2002 between 9:00 a.m. and 10:00 a.m.," or as general as "show me everything where George Bush was the topic." Other requests may include data regarding the date, time, language, and/or source of the desired information, or relevant words next to each other or within a certain distance of each other (similar to a typical database query).

**[0043]** Client 140 may process (e.g., convert) the request, if necessary, and issue the request to server 110 (act 405). For example, client 140 may establish communication with server 110 via network 150, using conventional techniques. Once communication has been established, client 140 may transmit the request to server 110.

**[0044]** Server 110 may formulate a query based on the request from client 140 and use the query to access metadata database 120. Server 110 may retrieve data (e.g., a transcription and metadata) relating to the desired information from metadata database 120 (act 410). Server 110 may then convert the data to an appropriate form, such as a Hyper Text Mark-up Language (HTML) document, and transmit the HTML document to client 140 for display in a standard web browser (acts 415 and 420). The HTML document may contain the transcription and metadata information, such as speaker identifiers, topics, and word time codes. In other implementations, server 110 may convert the data to another form or transmit the data unconverted to client 140.

**[0045]** Client 140 may present the HTML document to the user via a graphical user interface (GUI) (act 425). Fig. 5 is a diagram of an exemplary GUI 500 that client 140 may present to a user according to an implementation consistent with the principles of the invention. GUI 500 may be part of an interface of a standard Internet browser, such as Internet Explorer or Netscape Navigator, or any other browser that follows World Wide Web Consortium (W3C) specifications

for HTML. The information presented by GUI 500 in this example relates to an episode of a television news program (i.e., ABC's World News Tonight from January 31, 1998).

**[0046]** GUI 500 may include a speaker section 510, a transcription section 520, and a topics section 530. Speaker section 510 may identify boundaries between speakers, the gender of a speaker, and the name of a speaker (when known). In this way, speaker segments are clustered together over the entire episode to group together segments from the same speaker under the same label. In the example of Fig. 5, one speaker, Elizabeth Vargas, has been identified by name.

**[0047]** Transcription section 520 may include a transcription of the audio data. Transcription section 520 may identify named entities (i.e., people, places, and organizations) by highlighting them in some manner. For example, people, places, organizations may be identified using different colors. Topic section 530 may include topics relating to the transcription in transcription section 520. Each of the topics may describe the main themes of the episode and may constitute a very high-level summary of the content of the transcription, even though the exact words in the topic may not be included in the transcription.

**[0048]** GUI 500 may also include a play audio button 540 corresponding to an embedded media player, such as the RealPlayer media player available from RealNetworks, that permits the original audio corresponding to the transcription in transcription section 520 to be played back. As will be described below, the media player may access database of original media 130 to retrieve the original audio and present the audio to the user. For example, the media player may permit the audio corresponding to the transcription to be played.

**[0049]** GUI 500 may also include a product button 550. As will be described below, the product button 550 may be used by the user when the user desires to produce a translation of one or more portions of the document in transcription section 520.

**[0050]** Returning to Fig. 4, the user may read, skim, or browse the HTML document to determine whether the user would like to translate any portions of the document. To help the user make this determination, the user may play back the information in the HTML document in its original form (act 430). In this case, the user may highlight or otherwise identify a portion of the HTML document for which the user desires to obtain the original audio and select play button 540. For example, the user may use a computer mouse to highlight the desired portion. Alternatively, the user may simply identify a starting point from which the original audio is desired.

**[0051]** Fig. 6 is a diagram of GUI 500 that illustrates a user's request to play back an original media. The user highlights a portion of the HTML document at highlighted block 610. The user selects play button 540 to initiate the playback process.

**[0052]** Returning to Fig. 4, when the user selects play button 540, client 140 initiates the embedded media player. The media player may determine the portion identified by the user, such as highlighted block 610 (act 435). In particular, the media player may identify the time codes, corresponding to the beginning and ending (if applicable) of the identified portion, using the time codes in the HTML document.

**[0053]** The media player may then retrieve the desired portion of the original audio signal (act 440). The media player may use conventional techniques to pull that portion of the original audio from database of original media 130. For example, the media player may use the

beginning and ending time codes (e.g., 7:03 p.m. to 7:05 p.m.) when accessing database 130. The original audio from database 130 may stream back to the media player. The media player can then play the original audio for the user (act 445).

**[0054]** As the media player plays back the original audio, client 140 visually synchronizes the playback with the transcription in the HTML document (act 450). To facilitate this, the media player lets client 140 know as time passes in the playback of the original audio. Because the metadata of the HTML document includes time codes that identify exactly when each word in the transcription of the HTML document was spoken, client 140 knows precisely (possibly down to the millisecond) when to highlight (or otherwise visually distinguish) a word. Client 140 compares the times emitted by the media player with the time codes and highlights the appropriate words.

**[0055]** Fig. 7 is a diagram of GUI 500 that illustrates the synchronization of the HTML document to the playback of the original media. Client 140 visually distinguishes the word "american" in synchronism with the playback of the original audio by the media player, as shown at the highlighted block 710.

**[0056]** The user may be permitted to stop the playback at any time. The user may also be permitted to control the playback by, for example, fast forwarding, speeding it up, slowing it down, or backing it up so many seconds or so many words. The media player or the graphical user interface may present the user with a set of controls to permit the user to perform these functions. Alternatively, the user may use foot pedals to control the playback of the audio signal.

**[0057]** The user may also be permitted to alter the HTML document in some manner and save the altered document back in metadata database 120. For example, the user may be

permitted to highlight or comment on the document that the user, or another translator, may desire to later translate. Client 140, in this case, may send the altered document back to server 110 for storage in metadata database 120.

**[0058]** At some point, the user may identify this document or another document as containing one or more portions that the user desires to translate. Fig. 8 is a flowchart of exemplary processing for translating an audio signal according to an implementation consistent with the principles of the invention. Processing may begin with the user viewing a document presented by client 140 that corresponds to an audio signal that the user desires to translate.

**[0059]** In translating the audio signal, the user performs three separate tasks: selection 810, translation 820, and publication 830. During selection task 810, the user selects a portion of the audio signal to translate (act 812). For example, the user may highlight or otherwise identify a portion of the document that the user desires to translate and select product button 550. In one implementation, the user may use a computer mouse to highlight the desired portion. Alternatively, the user may simply identify a starting point from which the user desires to begin the translation.

**[0060]** Upon selection of product button 550, client 140 may send a message to server 110 to retrieve the portion of the document selected by the user (act 814). The message may include data specific to the portion (or range) selected by the user. In response to the message, server 110 may obtain the text relating to the selected portion (i.e., the transcription of the audio signal relating to the range selected by the user) and send a return message to client 140. The return message may include the text and metadata (e.g., time codes, named entities, etc.) relating to the selected portion with the Multipurpose Internet Mail Extension (MIME) type set to inform client

140 that the text is intended for a translation application. The translation application may be a word processing application, such as Microsoft Word or WordPerfect from Corel Corporation, or another type of application, such as an application that operates upon Java or HTML.

**[0061]** Upon receipt of the return message, client 140 may initiate the translation application and present the text to the user (acts 816 and 818). Fig. 9 is a diagram of an exemplary graphical user interface (GUI) 900 that client 140 may present to a user in an implementation consistent with the principles of the invention. GUI 900 may be associated with the translation application to aid the user in translating an audio signal.

**[0062]** GUI 900 may include translation section 910 and transcription section 920. Translation section 910 may include the area in which the user types in, or otherwise provides, a translation of the audio signal. Transcription section 920 may include the area in which the text (or transcription) of the portion of the audio signal to be translated is displayed. GUI 900 may also include several buttons, such as backup button 930, play/pause button 940, save product button 950, and configuration (config) button 960. The functions performed when buttons 930-950 are selected will be described below.

**[0063]** Returning to Fig. 8, during translation task 820, the user translates the selected portion of the audio signal. The user may begin by selecting play/pause button 940. In response, client 140 may initiate an embedded media player, such as the RealPlayer media player available from RealNetworks. The media player may identify the time codes corresponding to the beginning and ending (if applicable) of the selected portion.

**[0064]** The media player may then retrieve the corresponding portion of the original audio signal (act 822). The media player may use conventional techniques to pull that portion of the

original audio from database of original media 130. For example, the media player may use the beginning and ending time codes (e.g., 7:03 p.m. to 7:05 p.m.) when accessing database 130. The original audio from database 130 may stream back to the media player. The media player then plays the original audio for the user (act 824).

**[0065]** As the media player plays back the original audio, client 140 may visually synchronize the playback with the transcription in transcription section 920 (act 826). To facilitate this, the media player lets client 140 know as time passes in the playback of the original audio. Because the time codes identify exactly when each word in transcription section 920 was spoken, client 140 knows precisely (possibly down to the millisecond) when to highlight (or otherwise visually distinguish) a word. Client 140 compares the times emitted by the media player with the time codes and highlights the appropriate words.

**[0066]** As the audio plays, the user may type in, or otherwise provide, a translation of the audio signal (act 828). Fig. 10 is a diagram of GUI 900 that illustrates a user's translation of an audio signal. If the user wishes to stop the playback of the audio, the user may select play/pause button 940. Play/pause button 940 may be toggled to start and stop the playback of the audio signal. Typically during translation, the user will need to replay portions of the audio signal. To do this, the user may select backup button 930. Backup button 930 may cause the playback of the audio to rewind a predetermined amount of time or number of words. The amount of rewind may be user-determined via configuration button 960.

**[0067]** When configuration button 960 is selected, the user may be presented with a configuration window, such as window 1010. Window 1010 may present the user with a number of options. For example, the user may be prompted to provide the product (i.e., the translation)

with a name. The user may also be prompted to identify a location at which the product is to be published (or saved). The user may further be prompted to identify the amount of rewind for each selection of backup button 930. The amount of rewind may be specified in terms of seconds or the number of words. If a number of words are specified, client 140 may convert the number of words to seconds based on the time codes associated with the text in transcription section 920.

**[0068]** In an implementation consistent with the principles of the invention, the functions of play/pause button 940 and backup button 930 may be initiated via one or more foot pedals. For example, the user may press a foot pedal to start and stop the playback of the audio. The user may press another foot pedal to back up a predetermined amount of time or number of words. The same or other foot pedals may be used to fast forward, speed up, and/or slow down the play back of the audio. The foot pedals may free the user's hands for typing in the translation.

**[0069]** Returning to Fig. 8, during publication task 830, the user may publish the translation results (act 832). The user may select save product button 950, which may cause configuration window 1010 (Fig. 10) to be presented to the user. The user, via configuration window 1010, may be given the option of saving the translation results to any file, directory, or location. The user may save the results to a location where it will be useful and may be easily accessed by people who may be interested in the translation.

## CONCLUSION

**[0070]** Systems and methods consistent with the present invention provide mechanisms that aid a human in translating an audio signal to another language. The systems and methods provide improvements at all three stages of the translation process. For example, the systems and

methods provide a transcription of the audio signal to the translator. This helps the translator in selecting a segment of the audio signal to translate because it is faster to skim through text than it is to listen to an entire audio signal. The translator may also use search criteria to find relevant text. This makes it possible to easily monitor a very large number of audio sources.

**[0071]** The systems and methods also present a transcription of the audio signal on the same screen that the translator uses to provide the translation. The systems and methods visually synchronize the playback of the audio signal with the text in the transcription. This helps the translator in translating the audio signal. For example, this gives the translator two indications (audible and visual) of what a particular word might mean, which increases the speed of translation. More people can read a language and translate it than can translate audio alone.

**[0072]** The systems and methods also permit the translator to publish the translation results anywhere that is useful. This helps the translator in making the translation results available to those who would be interested in them.

**[0073]** The foregoing description of preferred embodiments of the present invention provides illustration and description, but is not intended to be exhaustive or to limit the invention to the precise form disclosed. Modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention.

**[0074]** For example, it has been disclosed that a media player retrieves the original audio when instructed by a user. In other implementations, the original audio may be transmitted to the user along with the translation of the audio and any associated metadata. In yet other implementations, more than the requested portion of the original media may be transmitted to the user in anticipation of its later request by the user.

**[0075]** It may also be possible to translate the audio signal or the transcription of the audio signal using automated techniques. In this case, the translation may be presented to a human translator, possibly along with the transcription and/or the original audio signal, to aid the translator in preparing an accurate translation of the audio signal.

**[0076]** Further, while aspects of the invention have been described as operating upon speech within an audio signal, these aspects may also operate upon speech contained with a video signal. Still, while aspects of the invention have been described in reference to a client-server configuration over a network, systems and methods for translating in a manner consistent with the present invention may also be implemented locally on a single computer.

**[0077]** No element, act, or instruction used in the description of the present application should be construed as critical or essential to the invention unless explicitly described as such. Also, as used herein, the article "a" is intended to include one or more items. Where only one item is intended, the term "one" or similar language is used. The scope of the invention is defined by the claims and their equivalents.